DOCUMENT RESUME

ED 266 766 IR 011 982

AUTHOR Merrill, John
TITLE Levels of Questioning and Forms of Feedback: Instructional Factors in Courseware Design.
PUB DATE Apr 85
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 31-April 4, 1985).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Cognitive Processes; *Computer Assisted Instruction; Computer Simulation; *Courseware; *Feedback; Hypothesis Testing; Individualized Instruction; *Instructional Design; Microcomputers; *Questioning Techniques; Research Methodology; Tables (Data); Time on Task
IDENTIFIERS *Instructional Effectiveness

ABSTRACT

High and low level questions as determined by a panel of evaluators were combined with corrective feedback and attribute isolation feedback to form four versions of a computer-based science lesson. The sample consisted of 154 high school chemistry students in a suburban high school. The primary hypothesis was that students who received high level questions and attribute isolation feedback would perform significantly higher on posttests than students who received the other combinations of questioning and feedback. The experiment used the three-factor, repeated measures design known as the two between-one within-subjects design defined by between-subjects variables A (levels of questioning) and B (feedback conditions), which cross, and by the within-subject factor C (scores on Posttest I and Posttest II) treated as a repeated measurements factor. Results indicate the primary hypothesis was not sustained, but the question-level by posttest interaction and the question-level and posttest main effects were statistically significant at the .05, .001, and .01 levels respectively. The question-level main effect and the question-level by posttest interaction provided verification of the efficacy of levels-of-questions as an instructional factor in courseware design. High level questions were more likely to cause deep cognitive processing and to enhance retention than low level questions, and should be included appropriately in the design of computer-based instructional systems. A four-page list of references is provided. (Author/JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
OERI
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official position or policy.

FD266766

LEVELS OF QUESTIONING AND FORMS OF FEEDBACK: INSTRUCTIONAL FACTORS IN COURSEWARE DESIGN

John Merrill
Ph.D. Candidate

Department of Educational T[illegible]y and Practice
College of Education
The Ohio State University

Paper presented at the 1985 Annual Meeting of the American Educational Research Association, Chicago.

IR011982

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
John A. Merrill
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

LEVELS OF QUESTIONING AND FORMS OF FEEDBACK: INSTRUCTIONAL FACTORS IN COURSEWARE DESIGN

BY

John Austin Merrill, Ph.D.

The Ohio State University, 1985

Professor Keith A. Hall, Adviser

High and low level questions as determined by a panel of evaluators were combined with corrective feedback and attribute isolation feedback to form four versions of a computer-based science lesson. The sample consisted of 154 high school chemistry students in a suburban high school. The primary hypothesis was that students who received high level questions and attribute isolation feedback would perform significantly higher on posttests (administered one day and one week after exposure to the experimental treatment) than students who received the other combinations of questioning and feedback. The experiment used the three-factor, repeated measures design known as the two between- one within-subjects design defined by between-subjects variables A (levels of questioning) and B (feedback conditions), which cross, and by the within-subjects factor C (scores on Posttest I and Posttest II) treated as a repeated measurements factor.

The primary hypothesis was not sustained but the question-level by posttest interaction and the question-level and posttest main

IR011982

IR011982

effects were statistically significant at the .05, .001, and .01 levels respectively. (The absence of the feedback-form main effect indicates that the potential benefits of either feedback form were not fully available to the students and suggests that students need the opportunity to encounter the attribute isolation feedback more frequently in the lesson). The question-level main effect and the question-level by posttest interaction provided verification of the efficacy of levels-of-questions as an instructional factor in courseware design. High level questions were more likely to cause deep cognitive processing and to enhance retention than low level questions, and should be included appropriately in the design of computer-based instructional systems.

Problem Statement

Over the past several years important trends in the research and development of instructional technology have been identified (Kemp, 1981; Meierhenry, 1981) which attest to a sustained commitment of educators to seek answers to the many questions raised by the presence of advanced technology within educational settings. This study seeks in particular to add to the growing body of knowledge on the design and development of computer-based instructional systems by examining specific instructional factors related to courseware design. The need for this study arises for three reasons: (a) data are critically insufficient on appropriate uses of computers in education (Kearsley, Hunter, & Seidel, 1983; Schloss, Schloss, & Cartwright, 1984); (b) research conducted by Hall, Comer, Merrill, and Wenig (1983) strongly suggests that "probably the single most important factor in producing courseware of a high educational quality is the nature of the questions which are presented to the learner" (p.10); and (c) the commission report, A Nation at Risk: The Imperative for Educational Reform , refers to rapid changes fostered by computer-based technology as a major contributor to the need for reforms in our educational system but fails to suggest how this same technology might be used as a basis for reform (Peter, 1983).

The purpose of this study was to gain insight regarding the principal elements of quality courseware as media of instruction in their own right, such that high-quality, adaptive courseware can be designed. The instructional factors of questioning and feedback were chosen as the independent variables. No efforts were made to compare student learning from conventional instruction with learning from computer-based instruction, although that has been a central concern in other studies (Taylor, 1982).

Research

Appropriate terminology for conceptualizing computer-based instructional systems varies throughout the literature. In this study, computer-based education (CBE) was used as a comprehensive term that encompassed computer-based tutorials, computer-managed instruction, and computer-based instructional simulations.

Research to date has shown that CBE offers the potential of consistent, high-quality individualized instruction that is less vulnerable to the normal differences of teacher preparedness, school environment, peer pressure (Hall, 1982), and teacher-student

BEST COPY AVAILABLE

interaction. For CBE to achieve full potential, much more needs to be known about the elements of questioning and feedback which can engage learners in high-level thinking.

Questioning. The role of questioning in instruction has been a significant concern among educators for well over a half-century. Piaget and Dewey (Newton, 1978) proposed intellectual development through questioning, and over the years educators have advocated the use of effective questions to stimulate thinking among learners (Burton, 1929; Hunkins, 1968; and Taba, 1967). The interest of educators in the use of questions is reflected in (a) the many classification systems for categorizing questions which emerged during the 60's (Gall, 1970) and in (b) the inclusion of these systems in teaching methodology textbooks in the 70's (Ryan, 1973).

Despite the level of interest, Gall (1970) concluded that in the fifty years prior to his research there had been no essential change in the types of questions teachers emphasized in the classroom. The work of Ryan (1973), O'Neill (1979), and Cook and Mayer (1983) suggests that the level of questions defines the level-of-learning resulting from the instruction, and therefore should be conducted purposefully and systematically. Bloom's taxonomy (1956) was chosen for the present study to help define the levels-of-learning (recall, comprehension, application, analysis, synthesis, and evaluation) required of the learner during the CBE lesson. As a taxonomy it has received broad acceptance by educators and is supported by substantial empirical validation (Madaus, Woods, & Nuttall, 1973; Seddon, 1978).

Prior use of Bloom's taxonomy and other similar taxonomies (Gagne, 1962; Merrill, 1971) has been restricted primarily to conventional instructional environments. While the use of such taxonomies may provide positive results in some individual classrooms, the research literature does not establish definitive outcomes which indicate higher achievement due to higher-level questioning (Strykowski, 1983). Much of the earlier data was collected through classroom observation of instructional events which were not replicable. Andre (1979) concluded that "observational research on teachers' in-class questions is not an appropriate vehicle for systematically examining the effects of higher-level questions on student learning" (p. 283). CBE now provides a means of verifying the principles developed by Merrill and Tennyson (1977) and of designing courseware that includes more precision in questioning.

Feedback. Feedback is integral to the questioning process if learners are going to have the opportunity to confirm, revise, or extend their levels of conceptual understanding. Research on the uses feedback has been dominated by behavioral psychologists, but an expanded conception of feedback has been assisted through the development of information processing models.

BEST COPY AVAILABLE

Concern for the proper uses of feedback has grown with the increasing sophistication of computer-based instructional delivery systems. The particular form that feedback assumes may reflect any of several schools of thought, but historically the influence of the behavioral sciences has been prevalent. The impact of the cognitive sciences is much harder to ascertain because research on the cognitive dimensions of feedback is not very extensive.

As more educational researchers became interested in feedback conditions the feedback literature began to reflect a greater diversity of perspectives and research designs. Some researchers focused on feedback in relation to academic self-concept (Joseph & Maguire, 1982), while others were concerned with the inability of feedback to alter a student's natural problem solving strategies (Steinberg, 1980). Most of the feedback literature includes the effect of feedback versus no-feedback conditions on test or task performance, delayed versus immediate feedback, some combination of no-feedback and delayed feedback, and response confidence. When viewing the majority of studies to date a great deal of controversy has revolved around the efficacy of delayed versus immediate feedback. However, most of these studies focused on verbal learning tasks and not tasks involved with discrimination learning or concept formation and acquisition.

The concept instruction research of Merrill and Tennyson (1977) is the earliest example of a formal comprehensive effort to explicate the uses of an instructional form of feedback from a cognitive perspective. Their work is based on several major propositions, one of which is the most relevant for an understanding of feedback as used in this study:

> Correct classification of newly encountered instances is more probable if during instruction using "example" presentations each instance is accompanied by attribute isolation and if during instruction using "practice" presentations a learner's response (either right or wrong) is followed by attribute isolation feedback.
> (Merrill & Tennyson, 1977, p. 203)

This proposition is supported by earlier studies conducted by Merrill and Tennyson (1971), Young (1972), and Tennyson, Steve, and Boutwell (1975), in which the addition of attribute isolation to the concept learning task yielded significant results on posttest performance. Young's study in particular is relevant because treatment groups which were presented with instances in "practice" (inquisitory) rather than "example" (expository) mode performed significantly better on posttest performance.

BEST COPY AVAILABLE

Merrill (1983) has developed an even more extensive role for feedback in his Component Display Theory (CDT). In CDT Merrill defines a set of primary and secondary presentation forms (or strategies). The primary presentation forms (i.e., generalities and/or instances presented in an inquisitory and/or expository mode) act as the major vehicles of instruction, and the secondary presentation forms serve as elaborations of the primary presentations. Merrill characterizes feedback as a secondary presentation strategy that facilitates student processing in support of the primary ones.

Feedback in its elaborative forms may vary from the presentation of the correct answer to a complete reworking of the problem with the student (Merrill, 1983). Although Merrill (1983) does not use the term "attribute isolation" within CDT in reference to feedback, he does emphasize the need for some degree of attention focusing feedback at appropriate points in the instruction to help indicate to a student how an alternative representation of an example corresponds to the original.

The use of "correct answer" feedback within CDT at first seems analogous to corrective feedback in a lesson that provides knowledge of results and perhaps a restatement of the concept to be learned. Corrective feedback (CF) has been consistently criticized as being incapable of producing "true concept learning" because it only serves as external reinforcement and does little to affect a learner's cognitive strategies (Inhelder, Sinclair, & Bovet, 1974). Cohen (1983) criticized CF for providing only minimal remediation and cited this factor as being "one of the greatest deficits that software programs exhibit" (p. 23). However, CDT does not prescribe feedback in isolation from the context of the rest of the instructional process. For example, "correct answer" feedback may also be accompanied by graphic or pictorial presentations of relationships or may demonstrate the application of a principle to a particular example.

More than 50 experimental studies concerned with aspects of CDT have been conducted by Merrill (1983) and his associates. "There is considerable evidence to support the recommendations for secondary presentation forms" (Merrill, 1983, p. 331). Hence, the hypotheses for the present study have roots in earlier as well as current concept instruction research.

Feedback is defined here as an instructional event that provides information to the student following his response. Feedback can function to correct and/or reinforce a learner's response such that prior knowledge is strengthened or new knowledge is acquired. Feedback external to the learner has taken many forms in the research literature (Kulhavy, 1977), but the following two types have been identified for the purposes of this study:

BEST COPY AVAILABLE

a) corrective feedback (CF) - informs learners of the correctness or incorrectness of their answers and provides the full text of the correct answer in response to a student's incorrect answer. The full text of the answer may consist of a single word, phrase, or short paragraph.

b) attribute isolation feedback (AIF) - informs learners of the correctness or incorrectness of their answers and isolates the attributes of the concepts being studied to improve further responses. Attribute isolation helps to focus attention on the critical and variable attributes of a concept (Merrill & Tennyson, 1977). As a form of feedback AIF appears to offer the most promise for the interactive nature of CBE.

Corrective feedback in the present study might appear to be nothing more than a "straw man," but in comparison to earlier feedback studies in experimental psychology very little current data has been collected on appropriate forms and uses of feedback from courseware based on concept learning principles. Both CF and AIF were included in the research design for this study to test the effects of each separately and in combination with levels of questioning.

Methodology

The experiment used the three-factor, repeated measures design known as the two between- one within-subjects design. This design is defined by between-subjects variables A (levels of questioning) and B (feedback conditions), which cross, and by the within-subjects factor C (scores on Posttest I and Posttest II) treated as a repeated measurements factor (Kennedy, 1978, p. 454).

Research Hypotheses

The following general hypotheses were tested: (a) there will be a significant interaction between form of feedback and level of questioning from Posttest I to Posttest II, (b) there will be a significant main effect for level of questioning, and (c) there will be a significant main effect for form of feedback.

The primary hypothesis of the study was that students who received high-level questions and attribute isolation feedback in a computer-based lesson would perform significantly higher on Posttest I and Posttest II than students who received either high-level questions and corrective feedback, low-level questions and corrective feedback, or low-level questions and attribute isolation feedback in the same lesson.

The above hypotheses were derived from the literature review which indicated that high-level questioning will cause students to engage in deeper cognitive processing than low-level questioning. The case for attribute isolation feedback is not extensively supported by the literature and needs further substantiation.

BEST COPY AVAILABLE

Independent Variables

The independent variables (levels of questions and forms of feedback) were combined to create four combinations of questioning and feedback which were integrated with lesson content to form four versions of a computer-based lesson. Table 1 depicts the four combinations of questioning and feedback used in the research design.

Table 1

Experimental Treatments: Feedback Forms By Question Levels

Feedback Forms	Question Levels: Low	Question Levels: High
Corrective Feedback (CF)	Lo/CF	Hi/CF
Attribute Isolation Feedback (AIF)	Lo/AIF	Hi/AIF

Dependent Variables

The two dependent variables were scores on Posttest I and scores on Posttest II. Posttest II was similar to Posttest I in order to measure any retention effect caused by the combinations of questioning and feedback. Posttest I was administered one day after exposure to the experimental treatment. Posttest II was administered one week after exposure. Students were not informed that they were to take Posttest II. Posttest I and Posttest II were treated as a repeated measurements factor within the research design (i.e., as a within-subjects factor). The two tests are often referred to subsequently as the posttests.

The content of the CBE lesson was based on the Xenograde System, which is a collection of very small imaginary systems similar in structure to an atom or the solar system (Arlin, 1973; Martin, 1981; Merrill, 1964). Xenograde terminology and concepts were similar to the subject matter in a typical high school physics or chemistry course. High school science teachers and graduate students majoring in instructional design evaluated the lesson's content.

BEST COPY AVAILABLE

Pilot testing and data collection occurred from October to December, 1984 in local high schools. At the data collection site junior chemistry students were randomly assigned within a given class to one of the four experimental treatments. Administration of the experimental treatments to the different classes occurred at various times throughout the school day and on two successive days. Potentially confounding factors such as time of day, mental fatigue, and hunger were distributed throughout all groups by random assignment of students to experimental conditions.

Research Results

Posttest Results

The sample (N=154) represented the number of students on whom complete posttest data was collected. The random assignment of the sample resulted in the assignment of 31 students to Lo/CF, 39 students to Hi/CF, 40 students to Lo/AIF, and 44 students to Hi/AIF. The distribution across class standings for each experimental group is shown with posttest means in Table 2.

Table 2

Composition of Experimental Groups by Class Standing by Treatment with Posttest Means

Experimental Group	Class Standing			Experimental Group Totals		Posttests	
	Soph.	Jr.	Sr.			I	II
Hi/AIF	8	34	2	44	(28.6%)	7.386	6.545
Hi/CF	3	32	4	39	(25.3%)	6.794	6.410
Lo/AIF	0	36	4	40	(26.0%)	5.725	5.850
Lo/CF	3	26	2	31	(20.1%)	6.096	6.096
Totals	14	128	12	154			
Percentage of Sample	9%	83%	8%		100.0%		

Note. The posttest means have not been adjusted for unequal n's and are slightly different from those means presented in Table 5.

The sophomores were advanced placement students, and the seniors were taking chemistry to complete their science credits for graduation. Both sophomores and seniors had been assigned earlier in the school year to the junior classes by the school's scheduling office. Prior to statistical analysis of the posttest data the researcher examined the individual Hi/AIF scores to determine if the distribution of sophomores, juniors, and seniors across treatment groups contributed to the Hi/AIF group doing better than the other groups. Analysis of the individual class means within Hi/AIF suggested that the presence of the eight advanced placement sophomores increased the overall Posttest I mean. This effect diminished in Posttest II as shown in Table 3.

Table 3

Posttest Means for Hi/AIF Treatment Group by Class Standing

Treatment Group	Posttests I		Posttests II		Combined Means	
Hi/AIF	Soph.	Jr./Sr.	Soph.	Jr./Sr.	Soph.	Jr./Sr.
	8.12	7.22	6.25	6.61	7.19	6.92

The combined mean for the Hi/AIF sophomores was virtually identical to the combined mean for the Hi/AIF juniors/seniors. The presence of the sophomores in Hi/AIF does not seem to have given that group an undue advantage. The absence of advanced placement sophomores in Lo/AIF may or may not have contributed to that group's lower scores. Therefore, the data obtained from the sophomores was retained in the study for further analysis.

Late posttest scores. Some posttests were administered one or two days late because of absenteeism or the instructor's schedule. Of the ten students (6%) who took Posttest I late, six scored above the group mean ($\bar{X}$=6.61, s=2.03), three were .29 of one standard deviation below the mean, and one was slightly more (1.28) than one standard deviation below the mean. Scores from the late Posttest I group were retained for analysis because most of the scores were above the total group mean.

BEST COPY AVAILABLE

Four students (2.5%) took Posttest II late, but no single student took both posttests late. Of the four students, one scored above the group mean ($\bar{X}$=5.97, s=2.29), one was .92 of one standard deviation below the mean, one was slightly more (1.23) than one standard deviation below the mean, and one was almost two (1.84) standard deviations below the mean. Scores from the late Posttest II group were retained for analysis because one was a perfect score, two were in the same standard deviation range as 17.4% of the total sample scores, and the fourth score, although low, was only .61 of a standard deviation below the score immediately preceding it.

Unweighted means analysis. The experimental groups contained unequal and slightly disproportionate n's at the conclusion of the data collection. The method of unweighted means was chosen as the most appropriate form of analysis (Kennedy, 1978, p. 291) since the unequal n's resulted from the random assignment process and subject mortality and not from the nature of the specific treatments. The analysis of variance summary is presented in Table 4, and the interaction and main effect means of the experimental groups are presented in Table 5.

Table 4

ANOVA Summary for Interaction and Main Effects Sources

Source	df	MS	F
Between Ss			
Question Level(A)	1	92.92440	17.26162***
Feedback Form(B)	1	.00959	.00178
(AB)	1	15.20261	2.82403
S/AB	150	5.38330	
Within Ss			
Posttest (C)	1	11.64879	6.87796**
(AC)	1	10.91805	6.44650*
(BC)	1	3.02558	1.78643
(ABC)	1	2.65896	1.56997
SC/AB	150	1.69364	

* p <.05
** p <.01
*** p <.001

BEST COPY AVAILABLE

Three main effects(A , B , C), three first order interaction effects (AB , AC , BC), and one second order interaction effect (ABC - the primary hypothesis of the study) were investigated. From Table 4 it can be observed that one interaction source and two main effects were statistically significant: the question-level by posttest interaction ($p < .05$), the question-level main effect ($p < .001$), and the posttest main effect ($p < .01$). The primary hypothesis of a question-level by feedback by posttest interaction was not sustained.

Table 5

Interaction and Main Effect Means from Posttests I and II by Question Level by Feedback Form

Treatment Group	n	Posttest I $\bar{X}$	Posttest I s.d.	Posttest II $\bar{X}$	Posttest II s.d.	Question Level Means Across Posttests
Hi/AIF	44	7.773	1.816	6.614	2.137	
Hi/CF	39	6.949	1.835	6.564	2.049	
(High)		7.361		6.589		6.975
Lo/AIF	40	5.650	1.902	5.625	1.462	
Lo/CF	31	6.097	1.739	6.097	2.006	
(Low)		5.874		5.861		5.867
Total	154	6.675	1.996	6.240	1.957	

Note. The posttest means have been adjusted for unequal n's and are slightly different from those means presented in Table 2. Individual means were rounded. The total means for each posttest were taken directly from the computer output to avoid rounding error.

Interaction and main effects summary. The question-level by posttest interaction was significant with a difference of .77 between the high-level question means from Posttest I to Posttest II. The low-level question means were virtually identical and may represent a baseline for the experimental groups. A graphic representation of the question-level by posttest interaction would reveal that it was ordinal.

The question-level main effect indicated that students in the high-level question treatments performed better than students in the low-level question treatments at the .001 level of significance. The question-level main effect existed irrespective of feedback conditions.

The posttest main effect was significant at the .01 level with a decrease of .39 of one point from Posttest I to Posttest II. This result suggested strong retention over the one-week interval.

Time-on-Task

Task completion times varied from 25 to 58 minutes over the four groups. The TOT averages rounded to the nearest minute were: Lo/CF = 36; Hi/CF = 40; Lo/AIF = 39; Hi/AIF = 39. The differences among these TOT averages were not large enough to warrant using TOT as a covariate. For example, the Hi/AIF condition did not cause a student to spend much more time in the lesson on the average than the Lo/CF condition. Also, the spread of task completion times prohibited using TOT as a blocking variable because the n's per cell would have been too small.

The Lo/CF condition was expected to take less time than Hi/CF because low-level questions and corrective feedback would require less presentation time and less student processing time. Lo/AIF and Hi/AIF conceivably could have taken much longer than Lo/CF or Hi/CF because of the time required to read and process the attribute isolation summaries. However, when students in the Lo/AIF and Hi/AIF groups were correct on the first try, as usually was the case, they did not encounter the attribute isolation summaries and could proceed through the lesson at an average pace similar to students in the Lo/CF and Hi/CF groups.

Instrument Reliability and Validity

Point biserial correlation coefficients, the Kuder-Richardson formula 20 (K-R 20), mean item difficulty and mean item discrimination indexes, standard deviations, and the standard error of measurement were computed to determine the reliability of individual items and the overall validity of Posttest I and Posttest II. The coefficients and indexes were computed on a sample of 161, although the unweighted means analysis used a sample of 154.(1) Each of these indicators is

1. The loss of seven students falls within the acceptable limits ($\leq 5\%$) for maintaining statistical power.

BEST COPY AVAILABLE

displayed in Table 6.

Table 6

Point Biserial Coefficient, Mean Item Difficulty Index, Mean Item Discrimination Index, Standard Deviation, and Standard Error for Posttest I and Posttest II

Coefficient	Posttest I	Posttest II
Point biserial	.552	.507
Indexes		
K-R 20	.581	.672
Item difficulty	.341	.403
Item discrimination	.506	.584
Standard deviation	2.030	2.290
Standard error	1.430	1.450

The averages of the point biserial coefficients, computed separately, indicated strong relationships between each posttest. The moderately strong K-R indexes indicated homogeneous functioning among the items and thus reflected the content validity of the tests. The item difficulty indexes indicated that Posttest II was slightly more difficult than Posttest I. The rearrangement of test items for Posttest II may have contributed to this, but the lower mean score for Posttest II probably resulted more from memory loss than from increased item difficulty. The mean item discrimination indexes indicated that each posttest contained very good items, while the standard error indexes further indicated the accuracy of the tests.

Summary

The posttest results provided the data source for determining the existence of interactions and main effects. Although the primary hypotheses was not sustained, statistical significance was obtained for the question-level by posttest interaction, the question-level main effect, and the posttest main effect.

BEST COPY AVAILABLE

Interpretation and Discussion of Results

The investigation of question-levels and feedback-forms as instructional factors in courseware design was implemented and the results analyzed according to the proposed research design. The primary hypothesis, the question-level by posttest interaction, and the question-level, posttest, and feedback-form main effects are discussed along with methodological problems.

The question-level by posttest interaction and the question-level and posttest main effects were statistically significant. The absence of a statistically significant feedback-form effect will also be discussed since the feedback conditions were central to this study.

Primary Hypothesis

It was postulated that students who received high-level questions and attribute isolation feedback (AIF) in a computer-based lesson would perform significantly higher on Posttest I and Posttest II than students receiving either high-level questions and corrective feedback (CF), low-level questions and corrective feedback, or low-level questions and attribute isolation feedback in the same lesson.

The effects to be observed in the Lo/CF and Hi/AIF treatments were the main focus of the investigation because the greatest difference in means was expected here. The lack of attribute isolation feedback in Hi/CF may not have rendered the high-level questions ineffective, but to probe students with high-level questions and then to provide only corrective feedback is contrary to providing appropriate conditions for learning (Gagne & Briggs, 1979). The Lo/AIF treatment may have provided additional instruction in the form of attribute isolation feedback, but this may have been rendered ineffective by the presence of low-level questions, and would not form the basis of a strong instructional principle.

The primary hypothesis of a question-level by feedback by posttest interaction was probably not sustained for one or more of the following reasons: (a) the feedback forms were not encountered with enough frequency to impact student performance, (b) exposure to the experimental treatment was insufficient to demonstrate treatment effects, (c) treatment effects existed but the posttests were not sensitive enough to detect them, or (d) the subject sample was selected from a restricted range. The feedback-form effect will be discussed following The Posttest Main Effect , and the other three factors will be discussed under Methodological Problems .

BEST COPY AVAILABLE

The Question-Level by Posttest Interaction

The question-level by posttest interaction differs from the overall question-level main effect in that the interaction of one variable can be observed over time as a function of a second variable. In this case, question-level as a function of the posttest variable (i.e., the repeated measurements factor) was virtually unchanged from Posttest I to Posttest II, which indicated a strong retention effect created by both the high-level and low-level questioning treatments (see Table 5).

Students in the low-level question groups performed below the total group means for each posttest (6.675 and 6.240, respectively), but the similarity in posttest scores for Lo/CF and Lo/AIF suggested that the effect of the low-level questions was strong enough to maintain a retention effect over the one-week interval. Students in the high-level questioning groups showed a slight but insignificant decrease across posttest scores which further indicated the strength of retention caused by the question-level variable.

The Question-Level Main Effect

The question-level main effect indicated that students in the high-level question treatments performed better than students in the low-level question treatments at the .001 level of significance. This effect is encouraging because the high-level treatment contained some low-level questions for the purpose of concept presentation at the beginning of the lesson. (A transition to high-level questions was made after the fourth question). The level of cognition required for low-level concept learning is above that for associative learning, but students presented with only low-level questions are still not afforded the same opportunity to develop the skills to function successfully with high-level questions. The fact that the high-level treatment also had low-level questions diminished the number of high-level questions in the rest of the treatment and hence the opportunity for higher cognitive processing. A highly significant ($p < .001$) question-level main effect was still achieved. The question-level main effect result partially supported the research on questioning conducted by Hall, et al. (1983) and assumed added importance vis-a-vis the absence of the feedback-form main effect.

The difference in combined means between the high and low groups in Table 5 is important given the use of ten-item posttests and the average treatment exposure time of 38.5 minutes. A greater difference in means might have occurred if the high-level question treatment contained high-level questions exclusively or if expanded content material and an increased number of high-level questions were presented over a longer period of time.

Writing questions at various levels from the same content material constrained courseware development at times. For example,

BEST COPY AVAILABLE

BEST COPY AVAILABLE

low-level questions were difficult to write for the graph interpretations in Lo/CF and Lo/AIF given the graph's complexity relative to material presented earlier in the lesson. These low-level questions had to cause the student to apply his knowledge of xenograde systems without causing analysis of the graph. Increased diversity in the available lesson concepts might have provided more flexibility in writing the high-level questions without running the risk of writing trivial low-level questions based on the same concepts.

The Posttest Main Effect

The posttest main effect indicated strong retention for the one-week interval evident in the similarity of the posttests' means (see Table 5). Reasons for the presence of this effect may be threefold: (a) the structure of the computer-based lesson design facilitated concept acquisition, (b) the high-level questions caused deeper cognitive processing than the low-level questions, and/or (c) the posttests were identical in content although not identical in structure.

The structure of the lesson with its concept definitions, critical and variable attributes, examples, and accompanying feedback provided a framework in which students could process the lesson concepts. The degree to which the concepts were acquired and retained was a function of treatment effect.

The high-level questions were intended to cause deep cognitive processing within each student. The question-level main effect and the question-level by posttest interaction indicated that this processing occurred.

The fact that Posttest II was identical in content to Posttest I may have facilitated the recall of information. The sequence of test items and the order of response options within other items were rearranged to diminish the occurrence of simple recall and to yield a more accurate indicator of retention.

Analysis of the posttest main effect was based on a comparison of the combined high-question and low-question level means from Posttest I to Posttest II. The posttest main effect was significant at the .01 level and it is probable that the question-level main effect and question-level by posttest interaction contributed to the achievement of the posttest main effect.

The Feedback-Form Main Effect

The absence of the feedback-form main effect indicates that the potential benefits of either feedback form were not fully available to the students. A feedback-form main effect might have achieved statistical significance if students in the AIF groups had encountered the attribute isolation feedback more frequently. The attribute

BEST COPY AVAILABLE

isolation feedback was structured to provide concept analysis that would enhance a student's grasp of a given concept and to increase performance on posttest material. Subjects received the attribute isolation feedback after two incorrect responses. Data from a record-keeping utility suggested that students in the Lo/AIF and Hi/AIF groups may have received AIF less than one third of the time they were interacting with the lesson. Increased complexity in content and/or in question level may have caused more frequent AIF feedback.

The critical attribute (CA) summaries received after the first correct response in Lo/AIF and Hi/AIF were designed to highlight the concept attributes but not to the level of detail present in the attribute isolation feedback. The CA summaries may have provided confirmation but little, if any, new information to subjects highly confident about their choice of a correct answer. If this were the case, then the CA summary as part of the AIF condition may have had no more impact than a CF condition. A comparison of AIF means ($\bar{X}$=6.416) and CF means ($\bar{X}$=6.427) across posttests (i.e., the feedback-form main effect) revealed no difference in the relative impact of the two feedback conditions.

Methodological Problems

Forms of feedback was one of the independent variables but a feedback-form main effect was not achieved. Significance for this main effect and other interaction sources may have occurred with increased exposure to the experimental treatments, more sensitive posttests, or a less restricted sample range.

Exposure time. The times for lesson completion varied from 25 to 58 minutes over the four groups resulting in an average of 38.5 minutes to complete the lesson. The lesson material was designed as an hour of instruction to ensure adequate time for concept development and acquisition by the subjects. The fact that the questioning variable produced significant results despite an average single treatment exposure of just over one half hour enhances the importance of appropriate question-levels as an instructional design factor.

Treatment conditions could be modified if the benchmark of one hour is to be maintained in future research on these hypotheses. For example, the lesson for this study contained twenty questions. A more highly developed content and increased complexity of questions might have created divergent conditions capable of producing more significant effects.

Posttests. The posttests each contained ten items which may have been insufficient to detect a treatment effect. Short tests may contribute to measurement error because they do not contain enough items to capture variability among the observations (Glass, 1978,

BEST COPY AVAILABLE

p. 138). Future research on these hypotheses should use a lengthened set of posttests and contain a greater variety of content material.

Sample range. The two classifiation criteria used in selecting the sample for the study specified that the students be currently enrolled in high school and have a science background, preferably in chemistry or physics. No special selection criteria such as IQ were used to group students, so randomization of ability throughout all treatments was assumed. The chosen data collection site was a predominantly white suburban high school in an area with a middle to high socio-economic status. The homogeneity of the student population may have produced a restricted sample range for the experimental treatments in this study. Potential consequences of a restricted sample are reduced variability of observations (Glass, 1978) and an underestimation of an instrument's ability to measure (Popham, 1981). Both consequences are interrelated and may have contributed to a potential inability of the posttests to detect treatment effects.

Conclusions

The attainment of a significant question-level main effect provided verification of the efficacy of levels-of-questions as an instructional factor in courseware design. The question-level main effect was particularly important given the relatively brief treatment exposure time and the short posttests. High-level questions were more likely to cause deep cognitive processing and to enhance retention than low-level questions, and should be included appropriately in the design of computer-based instructional systems.

Recommendations for Future Research

The present study focused on the instructional factors of questioning and feedback with the intention of demonstrating the efficacy of those factors for the design of high-quality courseware. Future research endeavors for testing this study's hypotheses could be based on the same research design or on an extended model of the same design. Several possibilities exist with the present design. For example, the present study could be replicated with a younger age group (i.e., a middle school population) that would find the xenograde concepts more challenging. An increased perception of content complexity among younger students might increase the error rate and thus cause students to receive detailed feedback more frequently. The increased reliance on feedback might then result in a significant feedback-form main effect. Whether a younger subject sample was selected or a similar high school age group was used again, a less restricted sample range would be sought to increase the variability of observations. Programming changes could also be made to provide the complete attribute isolation summaries for correct responses. In any of these instances the goals would be to replicate the strength of the question-level main effect, to support the primary hypothesis, and to significantly increase the reliance on feedback.

BEST COPY AVAILABLE

The same research design could also be used with other subject matter - perhaps content that is more complex and thus amenable to the writing of more high-level questions. An increase in the number of high-level questions would necessitate lengthening the lesson or dividing it into two or more lessons. The latter option might be more desirable for data collection over a longer period of time covering a greater diversity of concepts. Significant research results on questioning and feedback based on cumulative data would strengthen the case for the inclusion of those factors in courseware design.

A logical extension of the two between- one-within subjects design is the three between- one-within subjects design. In the extended model the additional factor could be two levels of complexity of the same content or a comparison of highly structured content versus less structured content from unrelated disciplines. In either case, the present research design would be replicated <u>within</u> each level of the new factor.

Extending the present design to include two levels of the same content would create some challenges for writing questions, but the research results might provide valuable insight on matching content structure with appropriate question levels and feedback forms. The second option for an extended model could yield important results for determining which types of content are the most suitable for delivery via a computer-based format.

The present study examined the effects of question levels and feedback forms on the cognitive processing of individual learners, and brought further attention to the critical need for quality courseware. Additional research on instructional factors in courseware design is necessary to substantially contribute to our knowledge base in this area.

BEST COPY AVAILABLE

LIST OF REFERENCES

Andre, T. Does answering higher-level questions while reading facilitate productive learning? Review of Educational Research, 1979, 49 (2), 280-318.

Arlin, M. Learning rate and learning rate variance under mastery learning conditions. Unpublished doctoral dissertation, University of Chicago, 1973.

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. Taxonomy of educational objectives. Handbook 1: Cognitive domain. New York: Longmans, Green, 1956.

Burton, W. H. The nature and direction of learning. New York: Appleton, 1929.

Cohen, V. B. What is instructionally effective microcomputer software? Viewpoints in Teaching and Learning, 1983, 59 (2), 13-27.

Cook, L. K., & Mayer, R. E. Reading strategies training for meaningful learning from prose. In M. Pressley, & J. R. Levin (Eds.), Cognitive strategy research: Educational applications. New York: Springer-Verlag, 1983.

Gagne, R. M. The acquisition of knowledge. Psychological Review, 1962, 69 (4), 355-365.

Gagne, R. M., & Briggs, L. J. Principles of instructional design. New York: Holt, Rinehart, and Winston, 1979.

Gall, M. D. The use of questions in teaching. Review of Educational Research, 1970, 40 (5), 707-721.

Glass, G. V., & Hopkins, K. D. Basic statistics for the behavioral sciences. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1978.

Hall, K. A. Computer-based education. In H. E. Mitzell (Ed.), J. H. Best, & W. Rabinowitz (Assoc. Eds.), Encyclopedia of educational research (5th ed., vol. 1). New York: The Free Press, 1982.

BEST COPY AVAILABLE

Hall, Keith A., Comer, R., Merrill, J. A., & Wenig, G. Taxonomy of instructional strategies (vols. 1, 2 & 3). Dublin, OH: Instructional Design Associates, 1983.

Hunkins, F. P. The influence of analysis and evaluation questions on achievement in sixth grade social studies. Educational Leadership , 1968, 25 , 326-331.

Inhelder, B., Sinclair, H., & Bovet, M. Learning and the development of cognition. Cambridge, MA: Harvard University Press, 1974.

Joseph, C. H., & Maguire, T. O. The interaction betwen time of feedback and academic self-concept in level of performance in arithmetic skill. Journal of Educational Research , 1982, 75 (6), 360-365.

Kearsley, G., Hunter, B., & Seidel, R. J. Two decades of computer based instruction projects: What have we learned? Technological Horizons in Education Journal , 1983, 10 (3, Pt. 1), 90-94; 10 (4, Pt. 2), 88-96.

Kemp, J. E. Significant current trends in educational technology: Some implications. In Brown, J. W., & Brown, S. N. (Eds.), Educational media yearbook. Littleton, CO: Libraries Unlimited, Inc., 1981.

Kennedy, J. J. An introduction to the design and analysis of experiments in education and psychology. Washington, D.C.: University Press of America, Inc., 1978.

Kulhavy, R. W. Feedback in written instruction. Review of Educational Research , 1977, 47 (2), 211-232.

Madaus, G. F., Woods, E. M., & Nuttall, R. L. A causal model analysis of Bloom's taxonomy. American Educational Research Journal , 1973, 10 , 253-262.

Martin, M. A. A study of cognitive styles and the learning resource selection process (Final Report). Philadelphia, PA: Research for Better Schools, Inc., 1981.

Meierhenry, W. C. Trends in research and development in instructional technology. In Brown, J. W., & Brown, S. N. (Eds.), Educational media yearbook. Littleton, CO: Libraries Unlimited, Inc., 1981.

Merrill, M. D. Transfer effects within a hierarchical learning task as a function of review and correction on successive parts (Tech. Rep. No. 5, NONR 3985(04)). Urbana-Champaign: University of Illinois, Training Research Laboratory, 1964.

BEST COPY AVAILABLE

Merrill, M. D. Necessary psychological conditions for defining instructional outcomes. Educational Technology , 1971, 11 (8), 34-39.

Merrill, M. D. Component display theory. In C. M. Reigeluth (Ed.), Instructional-design theories and models: An overview of their current status. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1983.

Merrill, M. D., & Tennyson, R. D. Attribute prompting variables in learning classroom concepts. Instructional Research and Development Working Paper No. 28 , Brigham Young University, 1971.

Merrill, M. D., & Tennyson, R. D. Teaching concepts: An instructional design guide. Englewood Cliffs, NJ: Educational Technology Publications, 1977.

Newton, B. T. Theoretical bases for higher cognitive questioning: An avenue to critical thinking. Education , 1978, 98 (3), 286-291.

O'Neill, R. The fine art of asking the right questions. Praxis Des Neusprachlichen Unterrichts , 1979, 26 (1), 30-37.

Peter, H. J. Computers and educational reform: An imaginary dialogue. AEDS Monitor , 1983, 22 (5 & 6), 29-31.

Popham, W. J. Modern educational measurement. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1981.

Ryan, F. L. Differentiated effects of levels of questioning on student achievement. The Journal of Experimental Education , 1973, 41 (3), 63-66.

Schloss, P. J., Schloss, C. N., & Cartwright, G. P. Efficacy of four ratios of questions and highlights to text in computer assisted instruction modules. Journal of Computer-Based Instruction , 1984, 11 (4), 103-106.

Seddon, G. M. The properties of Bloom's taxonomy of educational objectives: Cognitive domain. Review of Educational Research , 1978, 48 (2), 303-323.

Steinberg, E. R. Experience versus two kinds of feedback in CAI problem solving. Paper presented at American Educational Research Association Annual Meeting, Reston, Virginia, 1980.

Strykowski, B. F. Meta-analysis of the effects of teachers' use of

BEST COPY AVAILABLE

Strykowski, B. F. Meta-analysis of the effects of teachers' use of higher-order questions on student achievement. Paper presented at American Educational Research Association Annual Meeting, New Orleans, April, 1984.

Taba, H. Teachers handbook for elementary social studies. Reading, MA: Addison-Wesley, 1967.

Taylor, J. A. The MEDICI melodic dictation computer program: Its design, management, and effectiveness as compared to classroom melodic dictation. Journal of Computer-Based Instruction , 1982, 9 (2), 64-73.

Tennyson, R. D., Steve, M. W., & Boutwell, R. C. Instance sequence and analysis of instance attribute representation in concept acquisition. Journal of Educational Psychology , 1975, 67 , 821-827.

Young, J. I. The effects of review techniques and instance presentation on concept learning tasks. Unpublished doctoral dissertation, Brigham Young University, May 1972.

BEST COPY AVAILABLE